ChinaXiv合作期刊

国内科学数据开放共享主题的作者合作关系与合作研究内容分析*

■ 盛小平　孙倩倩
上海大学图书情报档案系　上海 200444

摘　要：[目的/意义]揭示国内科学数据开放共享主题的作者合作关系与合作研究内容，以便为更广泛的科学数据开放共享提供参考。[方法/过程]利用 2004 - 2021 年 CNKI 中科学数据开放共享主题的期刊论文与会议论文数据，统计作者合作的基本情况，并构建作者合作网络，运用社会网络方法分析作者合作整体网络、合作团体、合作研究内容。[结果/结论]国内科学数据开放共享主题的期刊论文与会议论文以作者合作研究为主；作者合作的整体网络密度较低，合作团体规模和合作次数都会影响作者在网络中的中心性；作者合作团体结构呈现双核型、网架型、桥连接型、星型、流线型 5 种模式；作者合作研究内容覆盖大数据、科学数据、开放共享、数据出版、数据管理、数据政策、数据安全、政府数据、数据治理 9 个方面。

关键词：科学数据　开放共享　作者合作　合作关系　社会网络
分类号：G203
DOI：10.13266/j.issn.0252 - 3116.2021.23.002

chinaXiv:202304.00418v1

随着数据密集型科学范式的发展，科学数据的开放共享逐步成为学界共识。与此同时，为降低科学成本、加速科学成果产出，科学人员开展合作研究的需求日益增加。目前学界主要对科学数据开放共享政策[1]、利益相关者各角色责任与参与动力[2]、服务创新与评价[3]、平台与应用[4]、运行与保障机制[5]、对策与建议[6]等方面进行了广泛探索，但鲜见相关成果论述国内科学数据开放共享主题的作者合作关系与合作研究内容。本文拟对此问题做初步分析，以进一步促进国内科学数据开放共享实践的发展。

1　数据来源与研究方法

本文获取的“科学数据开放共享”主题下的数据均来自 CNKI 数据库，文献类型限定为期刊、会议，因为学位论文是由单个作者完成的，不涉及作者之间的合作，所以本次研究没有将其作为数据来源。使用主题词“科学数据”或“研究数据”，与主题词“开放共享”组合进行高级检索（检索时间为 2021 年 9 月 26 日），共获得相关论文 1 016 篇，其中期刊论文 1 000 篇，会议论文 16 篇，去除匿名作者（无作者标识和本刊讯类）论文和重复论文，将作者署名为“课题组”的论文作者替换为课题组成员，最终得到 884 篇论文（期刊论文 868 篇，会议论文 16 篇），作为进一步分析的样本数据。

经文献检索发现，早在 2004 年，科学数据开放共享这一主题在国内就已经受到关注，不过，2004 - 2012 年该主题发文量均为个位数，2012 年后该主题发文篇数、合著篇数、发文作者数量均呈现增长趋势。本文分析的合作主要以合著体现，在 884 篇论文中，多作者合作完成的论文为 492 篇；作者数量共计达到 1 518 位，其中独立作者（未参与合作发文）占 321 位，合著者有 1 197位；在单篇论文合作中，以 2 - 3 位合作者为多数，7 人以上合作完成的论文数量很少，有 1 篇论文的合作者达 23 位（见表 1）。简而言之，该主题论文合作度为 1.72（合作度 = 一定时期内相关文献作者总数/一定时期内相关文献论文总数 × 100%），合作率为55.67%（合作率 = 一定时期内相关文献合作论文数/一定时期内相关文献论文总数 × 100%），合作者占该主题总作者中的比例为 78.85%。这说明，大部分国内科学数据开放共享主题的期刊与会议论文都属于合作型研究。

* 本文系国家社会科学基金项目“开放科学环境下的科学数据开放共享机制与对策研究”（项目编号：18ATQ007）研究成果之一。
作者简介：盛小平，教授，博士，博士生导师，E-mail：shengxp68@126.com；孙倩倩，硕士研究生。
收稿日期：2021 - 07 - 11　修回日期：2021 - 10 - 04　本文起止页码：13 - 21　本文责任编辑：王传清

ChinaXiv合作期刊

表 1 单篇论文作者人数对应的论文篇数

单篇论文作者人数	1	2	3	4	5	6	7	8	9	14	15	23
论文篇数	392	246	128	59	32	16	3	2	3	1	1	1

本文主要运用社会网络分析方法和 VOSviewer、UCINET 等工具,从作者合作整体网络分析、作者合作团体分析、作者合作研究内容分析 3 个方面解析国内科学数据开放共享主题的作者合作情况。

2 作者合作整体网络分析

将作者作为合作网络的节点,将其共同署名的合作关系作为边,且不以作者在单篇论文中的贡献大小区分权重比例,可以构建无向对称的国内科学数据开放共享主题下作者合作的整体网络(见图 1),其中 a、b、c 为拥有合作关系的作者数量排名前三位的团体,团体 a 由 102 人组成,团体 b 由 29 人组成,团体 c 由 28 人组成。作者合作的整体网络是反映作者合作关系的一个主要方式,其密度分析与中心性分析如下。

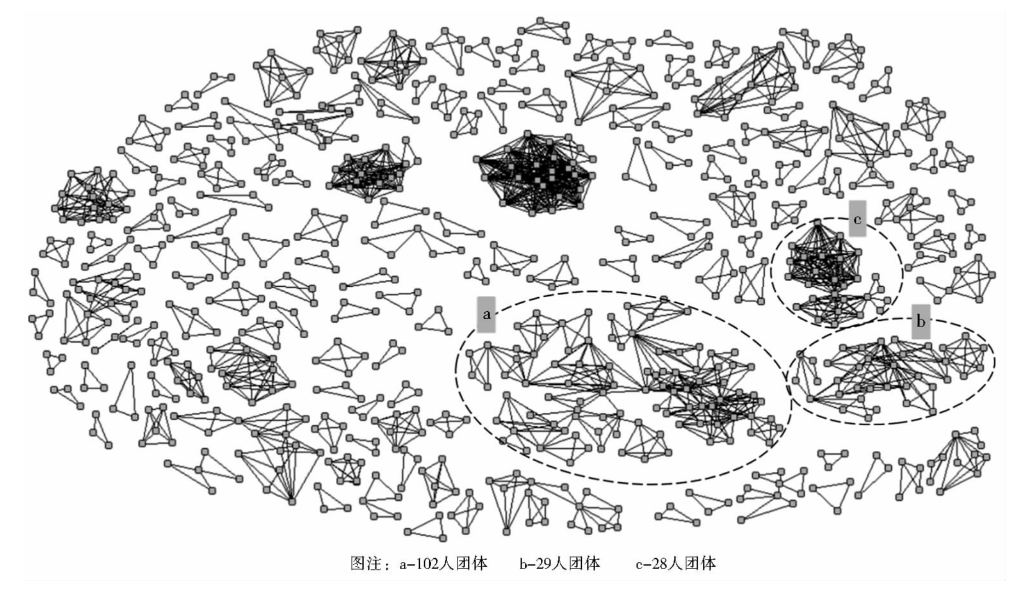


图 1 作者合作整体网络

chinaXiv:202304.00418v1

2.1 密度分析

网络密度反映了节点间关系的稀疏,密度值介于 0 – 1 之间,值越接近 1 则代表彼此间关系越紧密[7]。2004 – 2021 年,国内科学数据开放共享主题下的作者合作整体网络的密度仅为 0.004,网络中关系的标准差为 0.076,网络密度较低,整体网络具有分散性特征,连通性能较弱,即节点间的关系过于稀疏,表明作者之间的交流互动存在一定障碍。

2.2 中心性分析

中心性分析包含网络中节点的点度中心度、中间中心度、接近中心度的测量。在计算点度中心度时,由于一篇论文是由 23 人合作的,且其中的郝晋新、薛艳杰、崔辰州又与其他 3 位作者合作了一篇论文,所以,这 23 位作者的点度中心度都在 22 以上;参考普莱斯定律 $K=0.749\sqrt{N_{max}}$ 来筛选合作发文量高的核心作者,其中 N_{max} 为合作发文量最多的论文篇数。选取的样本数据中合作发文量最多的篇数为 17 篇,计算出 $K=3.09$,由此得出科学数据开放共享主题中参与合作的核心作者是合作发文篇数在 4 篇及以上的作者,共有 26 人(见表 2),而合作发文量排名前 3 位的作者黄如花、顾立平、盛小平的点度中心度分别为 15、19、13,低于上述 23 位作者。这是缘于该作者合作整体网络规模不大、分析样本数量不足够多而产生的一种特例,但点度中心度与合作团体规模大小有正向相关关系。

中间中心度数值越大,表明该节点处于越多节点经过的网络路径上[8]。在表 2 中,中间中心度数值最大的是石蕾,意味着她在整个网络中担任桥梁作用最大、掌握资源的能力最强。中间中心度排名前 26 位的作者中黄如花、黄雨婷属于 29 人团体,杨帆属于 28 人团体,其余 23 人均属于 102 人团体(见图 1)。但合作发文量排名前 26 位的盛小平、马海群、邢文明、温芳芳、胥婷、钱庆、于广军、李洋、吴思竹、张潇月、杨现民、朱扬勇并未出现在中间中心度排名前 26 位的作者中。这说明作者合作发表论文数量与作者中间中心度虽然

ChinaXiv合作期刊

盛小平，孙倩倩. 国内科学数据开放共享主题的作者合作关系与合作研究内容分析[J]. 图书情报工作,2021,65(23):13－21.

表 2　合作发文篇数与 3 种中心度排名前 26 的作者

作者	合作发文篇数	作者	点度中心度	作者	中间中心度	作者	接近中心度
黄如花	17	郝晋新	25	石蕾	2 696	石蕾	11 211
顾立平	13	崔辰州	25	高孟绪	1 838	高孟绪	11 229
盛小平	13	薛艳杰	25	王健	1 572	王瑞丹	11 243
马海群	9	杨帆	23	顾立平	1 365	王超	11 244
黎建辉	8	许允飞	22	胡良霖	1 119	胡良霖	11 246
温亮明	7	刘梁	22	陈志辉	737	徐波	11 250
高孟绪	6	杨丝丝	22	王瑞丹	732	顾立平	11 258
张丽丽	6	和兰	22	王超	728	黎建辉	11 259
王超	6	何勃亮	22	黎建辉	603	张丽丽	11 263
王瑞丹	6	米琳莹	22	张辉	486	朱艳华	11 274
王健	5	樊东卫	22	宋立荣	415	高瑜蔚	11 274
邢文明	5	杨涵溪	22	屈宝强	405	温亮明	11 282
温芳芳	5	李长华	22	樊景超	388	郑晓欢	11 286
胥婷	5	袁海波	22	赵华	388	王健	11 291
钱庆	5	刘峰	22	周园春	322	陈志辉	11 293
于广军	5	韩军	22	赵伟	294	杨静	11 304
李洋	5	陈肖	22	孙亮	294	周园春	11 306
吴思竹	5	肖健	22	张丽丽	291	李成赞	11 306
胡良霖	4	马捷	22	黄如花	291	张贵兰	11 309
张潇月	4	于策	22	李成赞	265	潘尧	11 309
杨现民	4	陶一寒	22	赫运涛	245	王玉洁	11 312
周园春	4	李珊珊	22	温亮明	212	王卷乐	11 312
李成赞	4	谌俊毅	22	董诚	198	王姝	11 316
朱扬勇	4	张海龙	22	杨帆	194	郭志斌	11 316
石蕾	4	顾立平	19	黄雨婷	180	刘佳	11 316
徐波	4	周园春	16	徐波	158	王丽娟	11 319

ChinaXiv:202304.00418v1

有正向关系，但不是简单的线性关系，作者中间中心度还与作者所在合作团体规模的大小有正向关系。

接近中心度数值越小，表明该节点与其他节点间的距离就越短，获取和传递信息更加容易且不必依赖他人。接近中心度排名前 26 位的数值都较大（见表 2），且接近中心度排名前 26 位的作者均属于最大规模的 102 人团体（见图 1），其中包括合作发文量排名前 26 位中的石蕾、高孟绪、王瑞丹、王超、胡良霖、徐波、顾立平、黎建辉、张丽丽、温亮明、王健、周园春、李成赞。但发文量排名前 26 位的余下 13 位作者没有出现在接近中心度排名前 26 位的作者中。这说明国内科学数据开放共享主题下的作者合作整体网络中的信息传递尚不够通畅，但在合作团体内则形成了较为良好的信息获取途径。因此，在国内科学数据开放共享主题的作者合作网络中，作者所在合作团体规模与合作次数，都会影响着其中间中心度与接近中心度。

3　作者合作团体分析

作者合作团体是反映作者合作关系的另一个主要参数。图 1 所示的作者合作整体网络结构分散，绝大部分团体相互独立，跨团体合作较少，因此，有必要进一步分析作者合作团体数量与结构。

3.1　作者合作团体数量分析

1 197 位参与合作的作者形成了 314 个合作团体。合作团体数量最多的是由 2 人组成的小团体，共 150 个；其次是由 3 人组成的小团体，共 84 个；有 9 个唯一的合作团体（见表 3）。最大规模的作者合作团体有 102 人，由 6 个小团体因“合作关系”而成（见图 2 左）；第二大规模的作者合作团体有 29 人，由 2 个小团体因“合作关系”而成（见图 2 右）。由此看来，国内科学数据开放共享研究主题已初步形成具有一定数量和不同规模的合作团体。

ChinaXiv合作期刊

表 3 合作团体数量与结构分布

团体人数	团体数量	细分的团体数量	团体网络结构类型	团体人数	团体数量	细分的团体数量	团体网络结构类型
2	150	150	双核型	9	4	2	桥连接型
3	84	79	网架型			1	星型
		5	流线型			1	网架型
4	40	35	网架型	11	2	2	桥连接型
		2	星型	12	1	1	桥连接型
		2	桥连接型	13	1	1	桥连接型
		1	流线型	14	1	1	桥连接型
5	9	4	网架型	15	2	2	桥连接型
		2	桥连接型	16	1	1	桥连接型
		2	星型	19	1	1	桥连接型
		1	流线型	26	1	1	桥连接型
6	7	5	网架型	28	1	1	桥连接型
		2	桥连接型	29	1	1	桥连接型
7	5	4	桥连接型	102	1	1	桥连接型
		1	星型				
8	2	1	网架型				
		1	桥连接型				

合计:双核型 150 个,网架型 125 个,桥连接型 26 个,星型 6 个,流线型 7 个

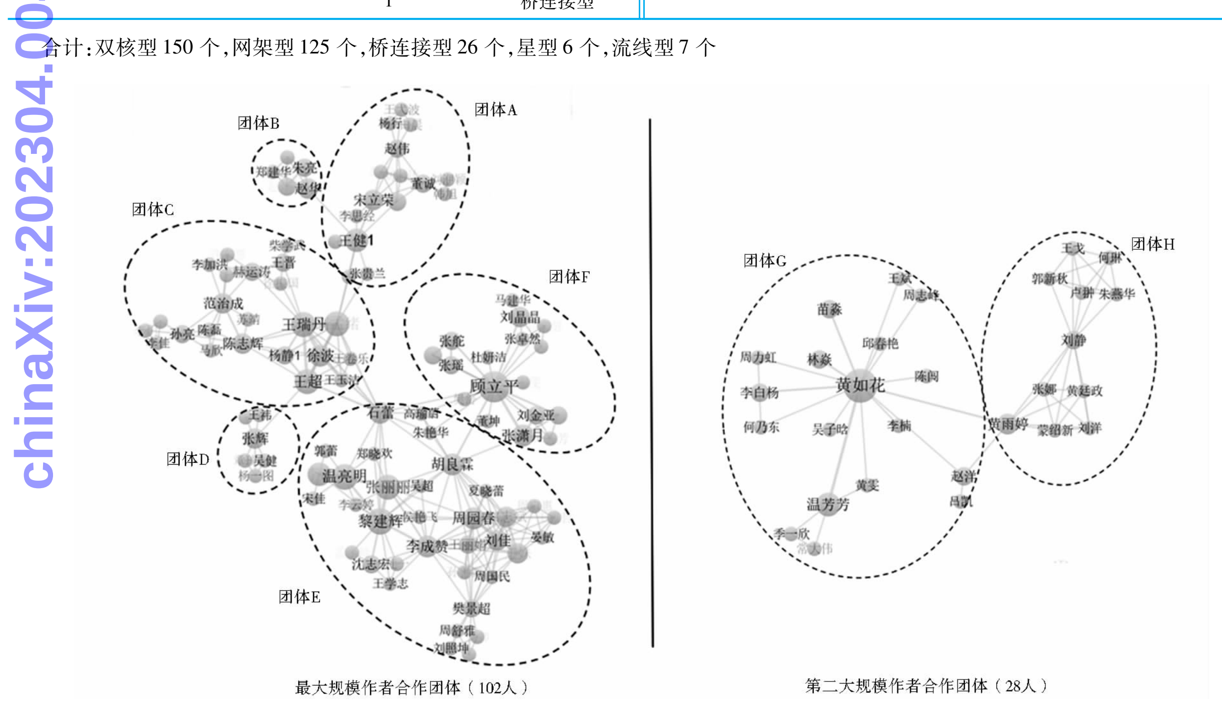


图 2 规模排名前两位的作者合作团体

3.2 作者合作团体结构分析

本文将科学数据开放共享主题下的作者合作网络中的拓扑结构类型归纳为双核型、网架型、桥连接型、星型和流线型 5 种[9],其中,双核型结构是指两人合作团体(见图 3(A));网架型结构是指任意两个节点之间都有联系的作者合作团体模式(见图 3(B));桥连接型结构是指需要通过桥节点将多个节点或合作团体连接起来的作者合作团体模式,表现为混合网络结构状态(见图 3(C));星型结构是指以一个节点为核心将分散的节点连接起来的作者合作团体模式(见图 3(D));流线型结构是指节点与节点以线性方式而非交叉方式形成合作关系的作者合作团体模式(见图 3(E))。在 5 种模式中,最多的是双核型结构(占 47.8%),其次是网架型结构(占 39.8%),流线型结构仅出现在 3-5 人的小型合作团体中,而 10 人以上的合作团体均呈现桥连接型结构(见表 3)。

chinaXiv:202304.00418v1

ChinaXiv合作期刊

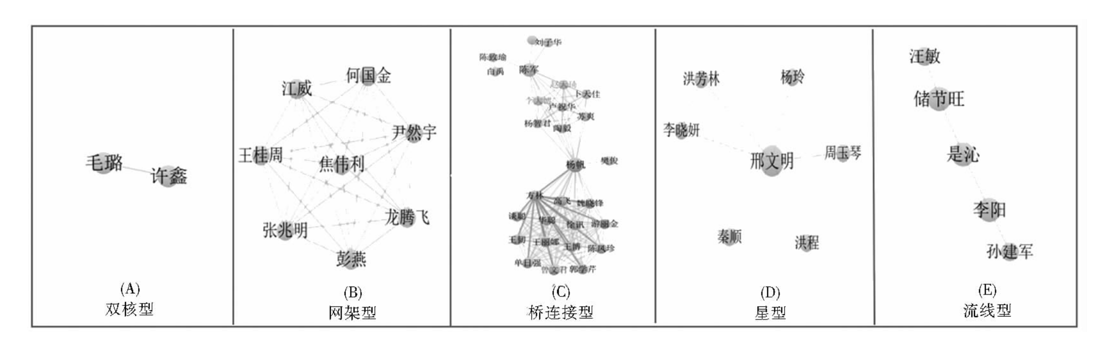


图 3　合作团体网络结构模型举例

4　作者合作研究内容分析

选择 492 篇合作论文进行作者合作研究的内容分析,从中提取出关键词 1 158 个,在对同义关键词如科学数据与研究数据、政府数据开放与开放政府数据等合并的基础上,构建关键词矩阵后导入 VOSviewer 中,选择出现 7 次以上的高频关键词进行聚类,生成高频关键词共现图谱(见图 4),主要包括大数据、科学数据、开放共享、数据出版、数据管理、数据政策、数据安全、政府数据、数据治理 9 个分主题,其合作研究内容如下:

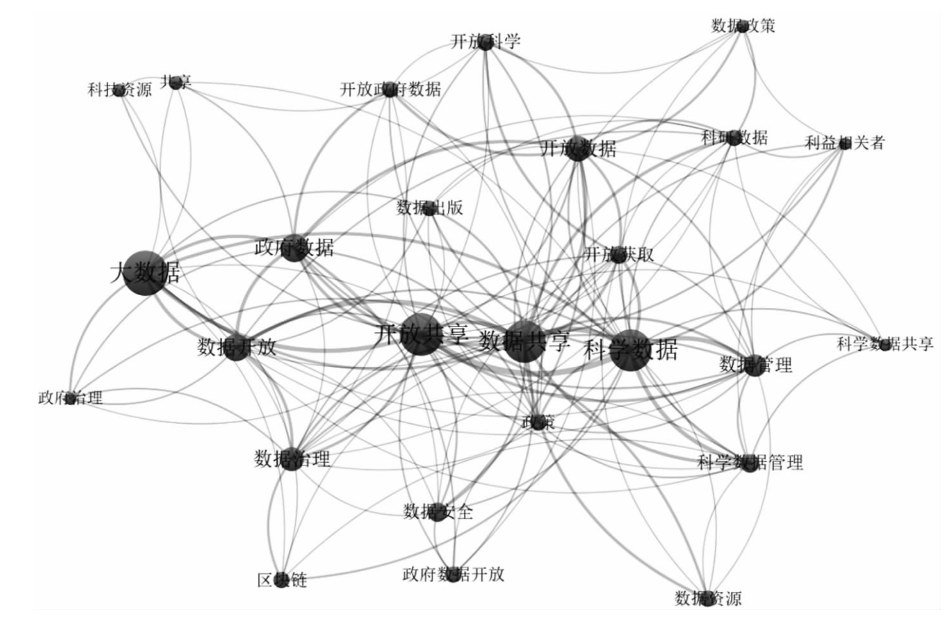


图 4　大于 7 次的高频关键词共现图谱

4.1　大数据

该主题集中于科学数据开放共享的大数据背景、大数据领域、大数据技术以及大数据对开放共享的要求等方面。大数据时代,数据的处理与应用方式均产生了巨大的变化,科学研究进入以数据密集型科学发现为主要特征的科学研究新范式,对科学数据的保存、管理和共享需求不断提升[10],同时在战略、公众等众多领域产生了相应的开放共享需求与多样的开放共享模式。人们特别关注健康医疗大数据、农业大数据、对地观测大数据、信用大数据、政务大数据等方面的科学数据资源现状和共享服务。然而,由于传统的数据处理软件无法在规定时间内处理各种类型的大数据,所以需要借助大数据技术,如大数据图像识别技术、视频录制技术、信息采集技术等,并结合物联网、云存储、人工智能等新一代信息和网络技术,共同助力科技资源共享,以实现对数据的集成、挖掘、展示及管理[11]。因此,大数据时代需要加强对科学数据的质量管理,确保开放共享科学数据的规范性、准确性、真实性和时效性;建立机构联盟大数据库,加强同类型机构之间的数据信息共享和互联互动[12],促进不同类型、不同学科领域科学数据的交叉融合与综合应用。

4.2　科学数据

该主题聚焦于科学数据的内涵与特点、现状与问题、管理政策、影响因素以及应对策略等方面。科学数

chinaXiv:202304.00418v1

ChinaXiv合作期刊

据是指科技活动中所产生的各类基础性、观测性的数据资料和相关信息,是科学成果最重要的信息载体,具有客观性、多样性、基础性、资源性、传递性、共享性、增值性等特点[13]。国内高校和研究机构陆续建立数据仓储,出版社相继推出数据期刊,商业公司也推出各种数据服务,科学数据集的数量呈指数增长态势[14]。但与国外相比,国内科学数据共享平台建设起步相对较晚,法律法规缺失,制度保障体系尚待完善;集成度有限,数据来源相对单一。国内外科学数据开放共享受到科学人员、政策、数据、技术、组织、平台、法律和资金8个方面影响[15]。为进一步推进科学数据开放共享,应从健全法律法规、政策和标准体系,完善管理规划和监管模式,利用新一代信息技术实现技术保障,增强内外合作、协同共建,加强人才队伍建设等方面打造开放共享体系[16]。

4.3 开放共享

该主题集中于科学数据开放共享的原则、共享模式、运行模式与实现途径、存在问题、关键机制与策略等方面。FAIR(可发现、可访问、可互操作、可重用)已经成为科学界广泛接受和认可的科学数据开放共享原则[17]。目前,国内外已有的科学数据共享模式可总结为以下5种:政策主导型、组织合作型、专题驱动型、数据出版型、数据集市型[18]。科学数据开放共享的运行模式通常包括政府主导下的公益性运行模式和市场主导下的商业化运行模式,其实现途径主要有数据知识库、数据期刊、学术期刊的数据说明文件3种[19]。然而,我国科学数据的开放共享尚属基础性、理论性探索阶段,仍缺少强有力的政策法规的支撑,与国际组织的业务沟通和项目合作偏少;科技数据资源分布不均,共享进度不均衡;科技资源共享平台建设繁杂,缺乏统一标准;传统的保守思想观念浓厚,制约共享积极性[20]。因此,需要通过有效的激励机制不断强化数据开放共享驱动力,通过长效评价机制不断完善数据开放共享模式,通过多级传播机制拓展开放数据传播渠道,以实现数据资源从开放到传播到再利用完整链路的良性循环。此外,我国亟待采取有效策略来促进科学数据开放共享,这包括构建科学数据开放共享体系,加强科学数据的保存、积累和分析挖掘,对科学数据权利进行合理界定和有效保护,逐步推动科学数据的分级分类开放,提升科学数据开放共享的国际合作实践[21],开发标准化数据产品来实施开放共享[22]等。

4.4 数据出版

该主题聚焦于科学数据出版的作用和要求、出版模式和出版流程、期刊数据出版政策、数据出版面临的问题和对策等方面。数据出版不仅可以推动数据开放共享,促进整个学术出版和科学研究的发展,而且可以利用数据开放共享解决数据知识产权问题,保障科学数据的质量,提高数据重用的价值[23]。科学数据出版需要达到科学数据标准化引用、唯一化识别、便捷化利用的要求[15]。数据出版模式包括独立的数据出版(即在数据存储库存储发布)、作为论文辅助资料的数据发布、以数据论文形式发布3种类型[24]。一般的数据出版流程包括数据提交、数据存储、数据评审、数据引用和数据评价5个环节[25]。目前,国际大型出版机构数据出版政策涉及数据提交、数据评审、数据开放与存储、数据保护4个方面,积极鼓励作者提交和共享数据;而国内期刊数据政策对数据出版的要求与说明还不成体系[26]。数据出版面临的主要问题是,数据出版与传统出版相比在树立学术地位和行业认可度等方面还有待提升,其还存在运营模式尚不够成熟[27]、数据存储设施建设不足、数据评审体系不健全、数据知识产权侵权严重、数据引用效果不佳等问题[28]。为应对这些挑战,我国科学数据出版应加快建立一批开放、规范、严格的质量控制机制、支持FAIR原则的数据存储库;探索新型出版模式,建立数据出版中的权益保护机制;规范数据引用;完善数据评审体系,保障科学数据质量;重视数据知识产权,推动数字资产管理[29]。

4.5 数据管理

该主题聚焦于数据管理的对象与原则、科学数据管理规划与办法、科学数据管理机构与管理服务、实践现状与发展建议等方面。数据管理的对象主要是大数据、政府数据、科学数据、企业数据[30],其原则主要包括数据价值量化、数据质量控制、元数据管理、计划管理、团队协作、动态管理、生命周期管理、风险管理、技术关联性、领导支持原则[31]。科学数据管理规划描述了将要编写的数据以及在整个生命周期中如何管理数据并使之可访问,主要包括项目预期产生的数据内容、类型、规模、质量、提交时间和最终汇交的科学数据管理机构名称等,有助于有计划地推进科学数据管理。《科学数据管理办法》为我国科学数据管理提供了宏观指导,其内容、落实现状、影响因素及推进策略等受到关注[32]。科学数据管理各级单位需要按照《科学数据管理办法》明确责任担当,协同参与科学数据管理实践;图书情报机构应当大力推广科学数据管理服务[33];数据管理机构应该广泛应用新技术,积极推动开放数据管理的成熟与发展。国内科学数据管理实践

chinaXiv:202304.00418v1

ChinaXiv合作期刊
盛小平，孙倩倩. 国内科学数据开放共享主题的作者合作关系与合作研究内容分析[J]. 图书情报工作,2021,65(23):13-21.

日趋成熟,但科学数据管理多集中于自然与工程科学,社会科学数据管理实践仍有很大进步空间[34]。可以采取以下措施来促进我国科学数据管理:制定操作性强的实施方案,落实数据共享办法;加强科学数据共享监管、知识产权保护,并对数据跨境管理予以研究和规制;加强国家科学数据中心建设和科学数据持续整合;提高数据管理意识,培养相关人才等[27]。

4.6 数据政策

该主题着重于国内外组织或机构数据政策调研与文本内容分析。许多国际组织制定了科学数据开放共享政策,这些政策比较注重科学数据开放共享的价值、知识产权及其保护、互操作和高质量、评估和强调利益相关者的责任担当等内容[35]。国际出版机构、科学教育机构、科学资助机构等也制定了相关科学数据政策,涉及呈缴政策、存储规范、监督机制与质量保障规范、开放程度、传播规范和利用方式等内容。例如,Springer Nature 制定了鼓励、强烈建议、部分强制、全部强制性 4 个不同级别的数据许可政策[36]。我国初步搭建起了从中央到地方科学数据管理与共享的政策体系,但地区开放数据政策建设的状况参差不齐,且地方政府的开放数据政策主要以意见、通知、方案等形式颁布,政策的约束力和强制力有待提升,政策落实仍需继续积累[37]。某些数据政策内容如数据权益如何划分、网络安全如何落实等都值得进一步研究和探讨[38]。

4.7 数据安全

该主题注重数据安全的法律法规、政策、技术、治理等方面。数据安全涉及相关法律和标准,也涉及数据加密、软硬件数据保护机制、备份、数据屏蔽和数据清除等技术[39]。在没有颁布《中华人民共和国数据安全法》之前,我国保障数据安全的相关法律主要是《中华人民共和国网络安全法》《中华人民共和国国家安全法》等[40]。目前我国数据安全政策着重于网络安全保障、个人隐私保护、数据跨境流动以及开放数据平台的安全保障等方面的部署[41],同时比较关注数据安全治理。在实施数据安全技术治理过程中,应充分利用区块链技术[42]、隐私增强技术、数据认证技术与构建基于数据联盟的国家科学数据中心[43]。不过维护数据安全不仅是一个技术问题,更是一个管理的问题[44]。因此,需要加强数据安全管理治理,包括建立科学数据分级分类标准与系统、实施数据保护影响评估、提高科学数据质量[15]。从长远来看,数据安全固然重要,但是数据开放共享是大势所趋,因此需要完善数据开放保障机制,在数据开放与数据安全之间找到均衡点[45]。

4.8 政府数据

该主题聚焦于政府数据开放的价值、政策与标准体系、问题与对策等方面。政府数据开放主要是指政府机构在法律法规范围内开放、共享其生产或拥有的按照一定标准规范组织过的数据集,可供企业、公民、研究者等自由使用,为社会创造价值[46]。政府数据开放具有政治、社会、经济和技术价值[47]。自 2015 年国务院印发《促进大数据发展行动纲要》以来,中央和地方政府制定了一系列数据开放政策,涉及数据基础设施政策、基于开放政府数据生命周期的数据资源管理政策、政府治理政策等[48]。国外政府数据开放的政策法规与标准[49]、合作模式与建设经验[50]也引起了国内学者的兴趣。政府数据开放共享需要关注政府数据及元数据、数据管理、数据安全和隐私、数据权益和用户参与等问题。为应对这些问题,需要进一步规范政府数据开放途径,强化政府作品著作权保护,完善地方政府数据开放平台功能,注重对我国政府数据开放标准、数据质量控制、用户利用方面的研究[51]。

4.9 数据治理

该主题聚焦于数据治理的定义、范围与类型、存在的问题、治理思路与建议等方面。数据治理是制定标准、规则、政策和实施监督以确保遵守数据管理最佳做法的过程[31]。在治理范围上,数据治理可分为全球数据治理、国家数据治理、政府数据治理和企业数据治理[52]。由于国内数据治理相关理论体系尚未健全,数据治理开展过程中遇到了许多挑战和阻碍,在制度设计、技术能力、人力资源、管理和数据等方面都存在一些问题[53],如大数据关键技术研发滞后、数据治理体系层次不清、数据主权保护权责模糊、数据生产要素价值实现困难、数据安全和个人隐私面临风险等[54]。为进一步实施数据治理,应提高数据质量,建立数据交易流通制度,加强数据人才培养和数据治理体系建设,设置数据治理指导委员会和数据治理咨询委员会[55]等。

5 结论

通过对国内科学数据开放共享主题的作者合作整体网络、作者合作团体数量与结构、作者合作研究内容的分析,可以发现:①2012 年后,国内科学数据开放共享主题下的发文篇数、合著篇数、作者数量均呈现增长趋势,且期刊论文和会议论文一半以上属于合作研究成果。②国内科学数据开放共享主题中相关作者合作的整体网络密度较低,分散性较强、连通性较弱,意味

chinaXiv:202304.00048v1

ChinaXiv合作期刊

着这些作者之间的科学数据共享和交流互动存在一定障碍;而作者所属于的合作团体规模与合作次数,都会影响着他们在作者合作整体网络中的中间中心度与接近中心度。③国内科学数据开放共享研究主题已形成由 1197 位合作者构成的 314 个合作小团体,其中包括 150 个 2 人小团体、84 个 3 人小团体和其他小团体,最大规模的小团体是由 102 人组成;这些合作团体在科学数据开放共享主题作者合作网络中呈现双核型、网架型、桥连接型、星型和流线型 5 种拓扑结构模式,其中,双核型结构和网架型结构是该合作网络中的主要模式。④国内科学数据开放共享作者合作研究内容主要覆盖大数据、科学数据、开放共享、数据出版、数据管理与治理、数据政策、数据安全、政府数据、数据治理 9 个分主题,形成了一系列相关理论观点。今后随着科学数据开放共享的进一步发展,将可能在科学研究中出现新型的开放共享合作关系与合作研究内容,这是下一步研究值得继续关注的问题。

参考文献:

[1] 马海群,蒲攀. 国内外开放数据政策研究现状分析及我国研究动向研判[J]. 中国图书馆学报, 2015,41(5):76-86.
[2] 盛小平,吴红. 科学数据开放共享活动中不同利益相关者动力分析[J]. 图书情报工作,2019, 63(17):40-50.
[3] 武彤. 基于数据生命周期的美国研究图书馆科学数据开放共享服务研究[J]. 图书与情报, 2019(1):135-144.
[4] 周宏虹,伍诗瑜. 我国科技信息资源共享平台建设现状[J]. 科技管理研究,2019,39(5): 174-178.
[5] 卫军朝,闫妍,王海彪. 科学数据开放政策保障体系的构建与应用[J]. 图书馆学研究,2017(22): 46-54.
[6] 张潇月,顾立平,胡良霖. 国内外开放科研数据重用困境解决措施述评[J]. 图书馆,2021(3): 80-89.
[7] 刘军. 社会网络分析导论[M]. 北京:社会科学文献出版社, 2004:100.
[8] 邱均平,王菲菲. 基于 SNA 的国内竞争情报领域作者合作关系研究[J]. 图书馆论坛,2010, 30(6):34-40.
[9] 刘蓓,袁毅,BOUTIN E. 社会网络分析法在论文合作网中的应用研究[J]. 情报学报,2008, 27(3):407-417.
[10] 赵瑞雪,赵华,朱亮. 国内外农业科学大数据建设与共享进展[J]. 农业大数据学报,2019,1(1): 24-37.
[11] 李佳,王宏起,李玥,等. 大数据时代区域创新服务平台间科技资源共享行为的演化博弈研究[J]. 情报科学,2018,36(1):38-44.
[12] 李社旺,刘丹婷. 大数据视域下我国体育智库建设的现实困境和路径转型[C]//第十一届全国体育科学大会论文摘要汇编. 北京:中国体育科学学会,2019:699-701.
[13] 马宁,刘召. 科学数据资源开放共享体系研究[J]. 中国科技资源导刊,2017,49(3):1-7.
[14] 罗鹏程,崔海媛,赵静茹. 基于 DataCite 的科学数据现状特征研究[J]. 图书情报知识,2019(3): 101-112.
[15] 盛小平,袁圆. 国内外科学数据开放共享影响因素研究综述[J]. 情报理论与实践, 2021, 44(8):173-179.
[16] 李亚京,马海群. 基于文献计量的我国科学数据研究综述[J]. 图书馆研究与工作,2017(12): 46-51.
[17] 宋佳,温亮明,李洋. 科学数据共享 FAIR 原则:背景、内容及实践[J]. 情报资料工作,2021, 42(1):57-68.
[18] 李善青,邢晓昭. 大数据背景下的科学数据共享模式研究[C]// Proceedings of 2019 3rd international conference on advanced education and management science. Hongkong:Science and Engineering Research Center,2019:394-399.
[19] 刘晶晶,马建华. 论科研数据开放共享的三种途径[J]. 情报杂志,2015,34(10):146-150.
[20] 张旺,程慧平. 科学数据开放共享策略机制及优化路径研究[J]. 情报杂志,2020,39(5): 154-161.
[21] 李思经,宋立荣,王健. 面向开放共享的科学数据出版:机遇、挑战与对策[J]. 中国科技期刊研究,2021,32(5):671-679.
[22] 兰方信,吴兴洋,许平. 地面气象观测月数据文件中降水量提取及产品开发[C]//第 34 届中国气象学会年会 S20 气象数据:深度应用和标准化论文集. 北京:中国气象学会,2017:49-54.
[23] 储节旺,汪敏. 美国科学数据开放共享策略及对我国的启示[J]. 情报理论与实践,2019, 42(8):153-158.
[24] 李成赞,张丽丽,侯艳飞,等. 科学大数据开放共享:模式与机制[J]. 情报理论与实践,2017,40(11):45-51.
[25] 杨晶,康琪,李哲. 推动科学数据开放共享的思考及启示[J]. 全球科技经济瞭望,2019,34(10): 37-43.
[26] 刘兹恒,涂志芳. 数据出版及其质量控制研究综述[J]. 图书馆论坛,2020,40(10):99-107.
[27] 屈宝强,宋立荣,王健. 开放共享视角下科学数据出版的发展趋势[J]. 中国科技期刊研究, 2019,30(4):329-335.
[28] 孔丽华,刁妍,张晓林. 数据出版的趋势、机制与挑战[J]. 中国科学基金,2019,33(3):237-245.
[29] 许洁,唐文辉,夏心悦. 面向实践的数据出版现状分析与对策探讨[J]. 中国科技期刊研究, 2020,31(11):1331-1337.
[30] 黄海瑛,肖兰,王冰洁,等. 数据管理的研究进展与热点:战略、角色与应用[J]. 图书情报知识, 2017(5):22-29.
[31] 盛小平,宋大成. 数据管理与数据治理的比较分析及其对制定科学数据开放共享政策的启示[J]. 图书情报工作,2020,64(22):4-10.
[32] 邢文明,洪芳林,李晓妍. 科学数据管理体系的二维视角——《科学数据管理办法》解读[J]. 图书情报工作,2019,63(23): 30-37.
[33] 邱春艳,黄如花. 近 3 年国际科学数据共享领域新进展[J]. 图书情报工作,2016,60(3):6-14.
[34] 张丽丽,温亮明,石蕾,等. 国内外科学数据管理与开放共享的最新进展[J]. 中国科学院院刊, 2018,33(8):774-782.
[35] 尤霞光,盛小平. 8 个国际组织科学数据开放共享政策的比较与特征分析[J]. 情报理论与实践, 2017,40(12):40-45.
[36] 刘凤红,彭琳. FAIR 原则背景下国际出版集团的数据政策和实践[J]. 中国科技期刊研究, 2021,32(2):173-179.

chinaXiv:202304.00418v1

ChinaXiv合作期刊

[37] 谭必勇,刘芮. 我国地方政府开放数据政策研究——以 15 个副省级城市为例[J]. 情报理论与实践,2018,41(11):51－56.
[38] 顾立平,包秦雯. 国外开放科研数据的激励政策[J]. 图书馆论坛,2019,39(8):138－145.
[39] 丁红发,孟秋晴,王祥,等. 面向数据生命周期的政府数据开放的数据安全与隐私保护对策分析[J]. 情报杂志,2019,38(7):151－159.
[40] 盛小平,郭道胜. 科学数据开放共享中的数据安全治理研究[J]. 图书情报工作,2020,64(22): 25－36.
[41] 黄如花,苗淼. 中国政府开放数据的安全保护对策[J]. 电子政务,2017(5):28－36.
[42] 高永昌,高敬,李德强. 区块链在人社数据资产及敏感信息安全保护中的应用实践[C]//2020 年"网络安全技术与应用创新"研讨会论文集. 北京:《信息网络安全》北京编辑部, 2020:121－124.
[43] 李善青,郑彦宁,邢晓昭,等. 科学数据共享的安全管理问题研究[J]. 中国科技资源导刊,2019, 51(3):11－17
[44] 马海群,张涛,李钟隽. 新冠疫情下政府数据开放与安全的系统动力学研究[J]. 现代情报, 2020,40(7):3－13.
[45] 盛小平,王毅. 利益相关者在科学数据开放共享中的责任与作用——基于国际组织科学数据开放共享政策的分析[J]. 图书情报工作,2019,63(17):31－39.
[46] 肖敏,郭秋萍,莫祖英. 政府数据开放发展历程及平台建设的差异分析——基于四个国家的调查[J]. 图书馆理论与实践,2019(3):38－43.
[47] 黄如花,何乃东,李白杨. 我国开放政府数据的价值体系构建[J]. 图书情报工作,2017,61(20): 6－11.
[48] 黄如花,温芳芳,黄雯. 我国政府数据开放共享政策体系构建[J]. 图书情报工作,2018,62(9): 5－13.
[49] 赵蓉英,梁志森,段培培. 英国政府数据开放共享的元数据标准——对 Data. gov. uk 的调研与启示[J]. 图书情报工作, 2016,60(19):31－39.
[50] 武琳,伍诗瑜. 欧洲开放政府数据合作模式与实现——跨地区共建共享典范[J]. 情报资料工作,2017(4):75－80.
[51] 马伍翠,刘文云,苏庆收,等. 我国地方政府数据开放现状分析及发展对策研究[J]. 数字图书馆论坛,2019(3):34－41.
[52] 杨琳,司萌萌,朱扬勇. 多层级数据治理体系框架构建思路探究[C]//第十五届(2020)中国管理学年会论文集. 北京:中国管理现代化研究会,2020:927－934.
[53] 杨晶,康琪,李哲. 美国《联邦数据战略与 2020 年行动计划》的分析及启示[J]. 情报杂志,2020,39(9):150－156.
[54] 郑跃平,甘祺璇,张采薇,等. 地方政府数据治理的现状与问题——基于 43 个政务热线部门的实证研究[J]. 电子政务, 2020(7):66－79.
[55] 尹国伟,聂凤英,杜绍明. 美国涉农数据在线开放共享情况[J]. 农业大数据学报,2020,2(3): 84－94.

作者贡献说明:

盛小平:设计研究思路,修订论文;

孙倩倩:收集、处理数据,撰写论文。

Analysis of Authors' Cooperative Relationships and Research Contents on the Topic of Open Sharing of Scientific Data in China

Sheng Xiaoping Sun Qianqian
School of Library, Information and Archives, Shanghai University, Shanghai 200444

Abstract: [**Purpose/significance**] In order to provide references for more extensive open sharing of scientific data, the authors' cooperative relationships and research contents on the topic of open sharing of scientific data in china are revealed. [**Method/process**] Based on the data of journal papers and conference papers on the subject of open sharing of scientific data in CNKI from 2004 to 2021, the basic situation of author cooperations was analyzed, and the author cooperation networks were constructed. The overall author cooperation network, cooperative groups and cooperative research contents were analyzed by using the social network method. [**Result/conclusion**] The journal papers and conference papers on the subject of open sharing of scientific data in China are mainly based on author cooperative researches. The overall network density of author cooperation is low, and the size and frequency of cooperation affect the centrality. The structures of authors' cooperative group present five modes of double core, grid, bridge, star and streamline. The contents of authors' cooperative researches cover 9 aspects, including big data, scientific data, open sharing, data publishing, data management, data policy, data security, government data and data governance.

Keywords: scientific data open sharing author cooperation cooperative relationship social network

chinaXiv:202304.00418v1